iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** VII **Month of publication:** July 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Collaborative Filtering Approach of Keyword Aware Service Recommendation for Big Data Applications on Map-Reduce

Pavan Shanbhag[#1], Asst Prof Leena Sakri [*2]
[#]*Department Of Information Science and Engineering, S.D.M.C.E.T, Dharwad*

Abstract— Recommender systems are shown as valuable tools for providing acceptable recommendations to users. Within the last decade, the number of customers shopping online, services and on-line data has grown quickly, yielding the Big Data analysis as a drawback for service recommender systems. Recommendations systems have been very popular in research community, where many approach to provide recommendations to users are suggested. Most of existing service recommender systems present an equivalent ratings and rankings of services to completely different users while not considering various users' preferences, and thus fails to fulfill users' personalized needs. Proposed Service recommendation generates a personalized service recommendation list taking into consideration of the preferences of the user and recommends appropriate service to the user effectively. Keywords are used to indicate users' preferences. Filtering rule is adopted to come up with more accurate recommendations. The proposed method is implemented on Hadoop, a widely used platform for distributed processing of Big Data across clusters of servers using the Map Reduce data processing paradigm. Thus it is scalable and efficient in Big Data environment.

Keywords— Recommendations, Preference, Big Data, Hadoop, Map Reduce.

I. INTRODUCTION

Every day the size of data is being getting bigger from quite a while now. From the dawn of time to less than a decade ago (till 2003 precisely), mankind generated about 5 Exabyte of data. In 2012, global data grew up 2.7 Zettabyte [8] i.e., roughly 500 times more data than all data ever generated prior to 2003. Cisco estimates that, global data on web will grow 3 times bigger i.e., approx. 4.8 Zettabyte by the end of 2015 [9] .One reason for data growing bigger is that, its continuously being generated by variety sources such as sensors, cc TV cameras, social media, .etc and by variety devices.

Much of that data such as - Videos, photos, comments on social forums, reviews on various websites and so on is unstructured, which means data isn't stored on traditional structured predefined tables. Moreover the data sources are arriving so fast that there is not even time to store it and apply analytics to it. That's why the traditional data management and analytics tools alone don't enable IT to store, manage, process and analyze Big Data. "It Refers to 'Data' whose size is beyond the ability of current technology to - process, manage and capture the data within fixed elapsed time". We live currently in Data world. Everywhere we see only data..Thus it is important to learn how to store this data as well as how to process it. The solution to such an issue moving progressively from giving equipment to provisioning more sensible programming arrangements. Big Data likewise conveys new opportunities and discriminating difficulties to industry and the academia like most Big Data applications, the Big Data tendency likewise postures overwhelming effects on service recommender techniques. With the developing number of options for services, effectively recommending services that users favored have turn into an imperative research issue. Service recommender frameworks have been indicated as important tools to help users manage services over-burden and give proper recommendations to them. Cases of such useful applications incorporate CDs, books, and different items now utilize recommender frameworks. Over the most recent decade, there has been much research done both in industry and the academia on growing new methodologies for service recommender techniques.

II. MOTIVATION

With the success of the Web 2.0, more and more companies capture large-scale information about their customers, providers, and operations. The rapid growth of the number of customers, services and other online information yields service recommender systems in "Big Data" environment, which poses critical challenges for service recommender systems. Moreover, in most existing service recommender systems, such as hotel reservation systems and restaurant guides, the ratings of services and the service

recommendation lists presented to users are the same. They have not considered users' different preferences, without meeting users' personalized requirements. Following is an example in hotel reservation system illustrating such a case. Example 1. Swati and Anand are respectively browsing a hotel reservation website to reserve a hotel in Hubli, Karnataka, India. But the rating and recommendations of hotels provided by the website to them are the same. Consider that there are 3 hotels in Hubli: AR, CI and D. D is convenient to airport and also close to shopping mall. AR is near to the railway station and the bus stand, this convenient for transportation. The breakfast and dinner is delicious at Clarks inn and also has a good view. According to the website AR is better than D and D is better than CI. However, Swati prefers a shopping mall near the hotel, while anand is concerned about good food. So AR may not be appropriate choice for them, and D and CI may be more relevant to Swati and Anand respectively. So the challenge is to provide them with appropriate recommendation list in "Big Data" environment.

III. RELATED WORK

There have been numerous recommender frameworks grew in both the educated community and industry. In [7] the creators propose a Bayesian inference based recommendation framework for with respect to online informal organizations. They demonstrate that the proposed Bayesian induction based recommendation is superior to the current trust-based proposals and is equivalent to recommendations generated by Collaborative Filtering approach. In[6] , author give an outline of the field of recommender frameworks and portray the present era of recommendation system. They additionally depict different restrictions of current service recommendation approach, and talk about conceivable augmentations that can enhance recommendation abilities and make recommender systems appropriate to a significantly more extensive scope of utilizations. Most existing service recommender frameworks are just taking into account a solitary numerical rating to speak to an services' utility all in all[4]. Indeed, assessing an service through various criteria and making into note of client input can help to make more viable recommendations for the clients. With the advancement of distributed computing software projects, for example, Apache Hadoop, Map Reduce, and Mahout, it gets to be conceivable to design and implement scalable recommendation method in "Big Data" environment. The authors of [2] design and implement a Collaborative Filtering approach on Hadoop. They tackle the scalability issue by dividing dataset into smaller size chunks. In any case, their strategy doesn't have positive scalability and efficiency if the measure of data grows. [3] presents a parallel user-profiling methodology taking into account folksonomy data and implements a scalable recommender system by utilizing Map Reduce and Cascading methods. Jin et al. [5] propose an large-scale video recommendation technique taking into account item based collaborative filtering approach. They implement their proposed approach in Qizmt, which is a .Net based Map-Reduce approach, in this manner their approach can work for large scale video sites. As a rule, contrasting and existing methods, the proposed method uses reviews of past users to get both, active user & previous user preferences and the candidate services list, which makes recommendation more precise.

IV. METHODOLOGY

The main steps of proposed system are depicted in Fig-1. The system takes as an input the users' preference. The reviews of previous users are available to the system and are stored on HDFS. As shown in the figure, the system can be broken down into three steps as follows:

A. Capturing the preferences of the user

In this step the preference keyword set of active and previous user is formed. Basically the preference of active user is taken as input to the system using keyword aware approach on fixed candidate keyword list. The preference keyword set of the previous user is computed using their reviews stored on HDFS and by using the same candidate keyword list.

Volume 3 Issue VII, July 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

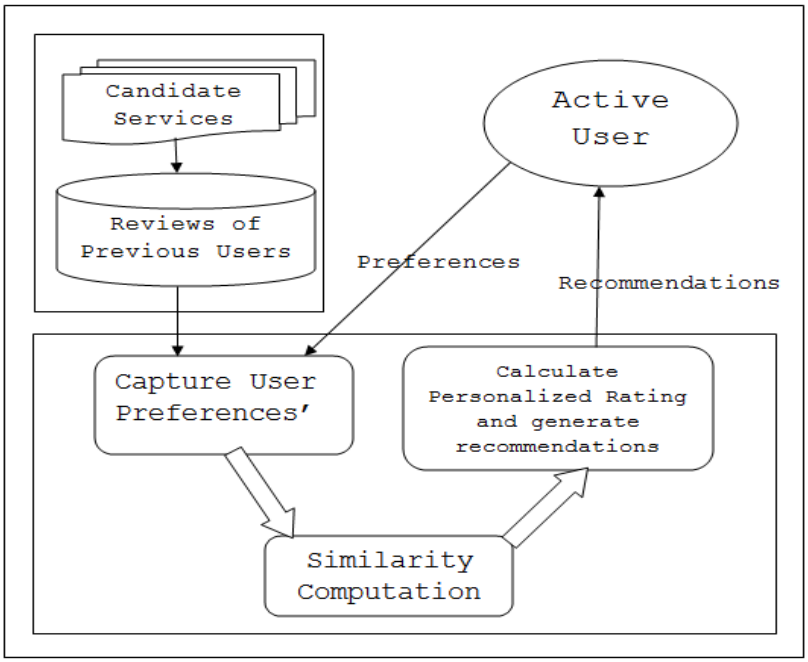


Fig-1: Main Steps in Proposed System.

B. Similarity computation

For all previous users individual preference keyword set is formed in previous step. In this step exact similarity of active user to each previous user is computed using a cosine-based approach shown below

Algorithm 1 Exact Similarity

Require: List containing keywords that denote preference of Active Users, PAU. List containing keywords that denote preference of previous user, PPU_j.

Ensure: The Similarity of PAU and PPU_j, $sim(PAU, PPU_j)$

1: termFreqVector[][]=TermFrequency(PPU, PAU_j)

2: termPPU=termFreqVector[0]

3: termPAU=termFreqVector[1]

4: $\text{Similarity}(PPU_j, PAU) = \frac{\overrightarrow{termPAU}.\overrightarrow{termPPU}}{Norm(termPAU)*Norm(termPPU)}$

5: **return** $Similarity(PPU_j, PAU)$

Cosine base algorithm takes into consideration the term frequencies of the terms in the two sets whose similarity is to be computed. The algorithm given below depicts how term frequencies are computed. The implantation of this is necessary in order to get similarity between the two sets using Algorithm1.

Algorithm 2 Term Frequency

Require: Two sets PPU and PAU containg preference keywords of previous user and active user respectively.

Ensure: termFrequencyVector containg term frequency of term $\in$ PPU and PAU.

1: Let $allTerms$ be the MAP of form <key,value>.
2: allTerms $= \emptyset$
3: **for** each keyword k $\in PPU$ **do**
4: **if** $!allTerm.contains(k)$ **then**
5: Insert $< k, 0X01 >$ into allTerm
6: **end if**
7: **end for**
8: **for** each keyword k $\in PAU$ **do**
9: **if** $!allTerm.contains(k)$ **then**
10: Insert $< k, 0X02 >$ into allTerm
11: **else** Insert $< k, 0X03 >$ into allTerm
12: **end if**
13: **end for**
14: $termPPU = \emptyset$ $termPAU = \emptyset$
15: **for** each value v $\in allTerm$ **do**
16: $termPPU = v \&\& 0X01$
17: $termPAU = v >> 1$
18: **end for**
19: **return** $termPPU, termPAU$

C. Compute personalized rating and generate recommendations for the user

Based on the similarity obtained in the previous step, i.e., similarity between active user and previous user, further filtering will be carried out. The personalized ratings will be calculated using the formula

$$pr_i = \bar{r} + k \sum_{PPU_j \in R} sim(PAU, PPK_j) * (r_j - \bar{r})$$

Where $\bar{r}$ average rating for the candidate service, PAU preference keyword set of active user, PPK_j is preference keyword set of previous user j , sim(PAU, PPK_j) is similarity computed in step 2 and r_j is rating of user j for candidate service. Using pr_i the user is recommended a list of hotels based on his/ her preference given as input to the system.

V. CONCLUSIONS

The proposed recommendation system uses, user based collaborative filtering approach to generate appropriate recommendation. The proposed system is implemented on Hadoop platform, and is applied to real time data set of a specific service domain (Hotel Reservation System). The system successfully generates personalized service recommendation list to the user as per the user preference. The performance of the system is observed by varying processor speed and RAM size for single node cluster.

VI. ACKNOWLEDGMENT

Leena Sakri is the co author of this paper. I would like to express deep sense of gratitude to Dr. Jagadeesh Pujari, Professor and Head of Department, Information Science and Engineering, SDMCET, Dharwad, for providing us an opportunity to carry out this work successfully.

REFERENCES

[1] Shunmei Meng, W. Dou, X. Zhang and J. Chen, "KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications," In IEEE Transactions on Parellel and Distributed Systems, vol. 25 , pp. 3221-3231, 2014.

[2] Z. D. Zhao, and M. S. Shang, "User-Based Collaborative Filtering Recommendation Algorithms on Hadoop," In the third International Workshop on Knowledge Discovery and Data Mining, pp. 478-481, 2010.

[3] H. Liang, J. Hogan, and Y. Xu, " Parallel User Profiling Based on Folksonomy for Large Scaled Recommender Systems: An Impli-mentation of Cascading MapReduce," In Proceedings of the IEEE International Conference on Data Mining Workshops, pp. 156-161, 2010.

[4] G. Adomavicius, and Y. Kwon, "New Recommendation Tech-niques for Multicriteria Rating Systems," IEEE Intelligent Systems, Vol. 22, No. 3, pp. 48-55, 2007.

[5] Y. Jin, M. Hu, H. Singh, D. Rule, M. Berlyant, and Z. Xie, "MyS-pace Video Recommendation with Map-Reduce on Qizmt," Pro-ceedings of the 2010 IEEE Fourth International Conference on Se-mantic Computing, pp.126-133, 2010.

[6] G. Adomavicius, and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the Stateof- the-Art and Possi-ble Extensions," IEEE Transactions on Knowledge and Data Engi-neering, Vol.17, No.6 pp. 734-749, 2005.

[7] X. Yang, Y. Guo, Y. Liu, "Bayesian-inference based recommenda-tion in online social networks," IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 4, pp. 642-651, 2013.

[8] What is Big Data?
http://ww-01.ibm.com/software/data/bigdata/what-is-big-data.html .

[9] Cisco global cloud index: Forecast and methodology 2013-2018 white paper.
http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html .

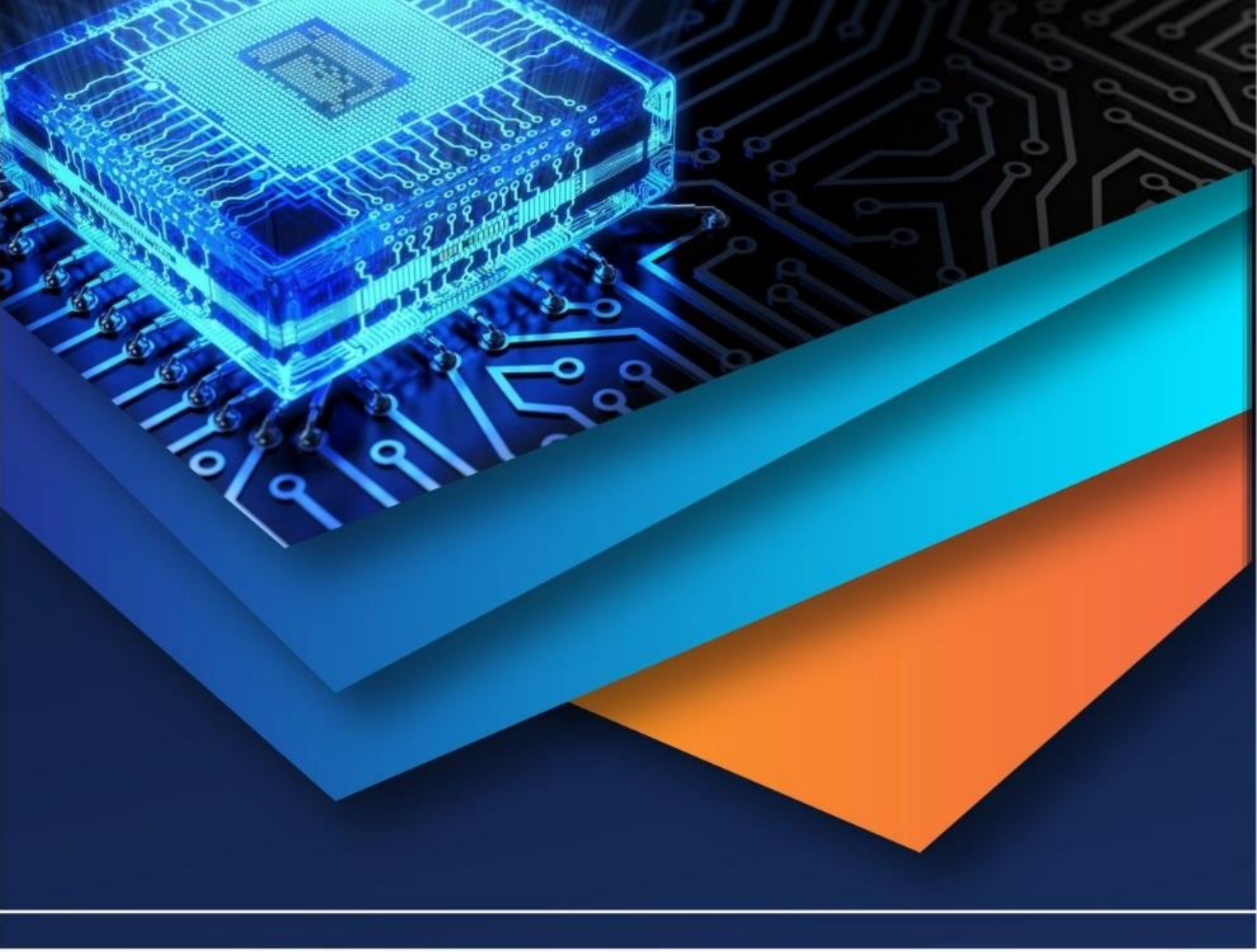
[10] W. Dou, X. Zhang, J. Liu, J. Chen, "HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applica-tions," IEEE Transactions on Parallel and Distributed Systems, 2013.

[11] Z. Luo, Y. Li and J. Yin, "Location: a feature for service selection in the era of big data," 2013 IEEE 20th International Conference on Web Service, pp. 515-522, 2013.

[12] J. Manyika, M. Chui, B. Brown, et al, "Big Data: The next frontier for innovation, competition, and productivity," 2011.

[13] M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for Sport Videos Based on User Audiovisual Consumption," IEEE Transactions on Multimedia, Vol. 14, No.6, pp. 1546-1557, 2013.

[14] Y. Chen, A. Cheng and W. Hsu, "Travel Recommendation by Min-ing People Attributes and Travel Group Types From Community-Contributed Photos". IEEE Transactions on Multimedia, Vol. 25, No.6, pp. 1283-1295, 2012.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY